Variance reduction in sample approximations of stochastic programs

Matti Koivu
Department of Management Science,
Helsinki School of Economics,
PL1210 00101 Helsinki, Finland

April 22, 2004

Abstract

This paper studies the use of randomized Quasi-Monte Carlo methods (RQMC) in sample approximations of stochastic programs. In high dimensional numerical integration, RQMC methods often substantially reduce the variance of sample approximations compared to MC. It seems thus natural to use RQMC methods in sample approximations of stochastic programs. It is shown, that RQMC methods produce epi-convergent approximations of the original problem. RQMC and MC methods are compared numerically in five different portfolio management models. In the tests, RQMC methods outperform MC sampling substantially reducing the sample variance and bias of optimal values in all the considered problems.

Keywords: Stochastic optimization, discretization, variance reduction techniques, randomized quasi-monte carlo methods, antithetic variates.

1 Introduction

Let Ξ be Borel subset of $\mathbb{R}^d$, and Σ the Borel σ-algebra on Ξ. Let P be a probability measure on (Ξ, Σ), and f an extended real-valued function on $\mathbb{R}^n \times \Xi$, such that $f(x, \cdot)$ is measurable for every $x \in \mathbb{R}^n$. This paper studies numerical solution through discretization of stochastic programs of the form

$$\underset{x \in \mathbb{R}^n}{minimize} \quad E^P f(x) := \int_\Xi f(x, \xi) P(d\xi), \tag{SP}$$

where the integral is interpreted as $+\infty$ when $f(x, \cdot) \notin L^1(\Xi, \Sigma, P)$. The decision variable x is not a function of ξ, so (SP) represents a *static* (one-stage) stochastic program. By allowing f to take on the value $+\infty$ we can incorporate constraints into the objective, which makes (SP) a very general model for optimal static decision making under uncertainty. Unlike most studies of stochastic programs, we do not assume the feasible set

$$\operatorname{dom} E^P f(x) = \{x \in \mathbb{R}^n \mid f(x, \cdot) \in L^1(\Xi, \Sigma, P)\}$$

to be known a priori. This is essential e.g. in stochastic programs without relatively complete recourse and in certain financial applications, where the determination of the feasible set is part of the problem rather than its statement; see Subsection 4.2.

A common approach to solving (SP), is to replace P by a finitely supported measure of the form

$$P^\nu = \sum_{i=1}^{\nu} p_i^\nu \delta_{\xi_i^\nu},$$

where $\delta_{\xi_i^\nu}$ denotes the unit mass located at ξ_i^ν. This yields

$$\underset{x\in\mathbb{R}^n}{minimize} \quad E^{P^\nu} f(x) := \sum_{i=1}^{\nu} p_i^\nu f(x, \xi_i^\nu), \qquad (SP^\nu)$$

which, is often easier to solve than (SP). In general, the aim is to choose P^ν so that (SP^ν) is a good approximation of (SP) and that the number ν of support points of P^ν is small enough to allow for numerical solution of (SP^ν). The simplest and the best-known method for numerical approximation of high-dimensional integrals is the Monte Carlo method (MC), i.e. random sampling. MC has also become the most popular method for constructing sample approximations of stochastic programs. However, in the literature of numerical integration there are many methods that usually perform better than MC in high-dimensional integration; see e.g. (Boyle et al., 1997, Jäckel, 2002). Quasi-Monte Carlo (QMC) methods can be seen as a deterministic counterpart to the MC method. They are designed to produce point sets that cover the d-dimensional unit hypercube as uniformly as possible. By suitable transformations QMC methods can be used to discretize many other probability distributions as well. They are just as easy to use as MC but they often result in faster convergence of the approximations thus allowing for smaller values of ν and cheaper computations.

L'Ecuyer and Lemieux (2002) review several QMC constructions and their randomizations that have been proposed to provide unbiased estimators and for error estimation. Randomizing QMC methods allows us to view them as variance reduction techniques. Randomized Quasi-Monte Carlo (RQMC) methods can be used just like MC in estimating confidence intervals and variances for sample approximations in numerical integration. RQMC often result in significant variance reduction with respect to MC. In this paper, we apply RQMC to stochastic optimization and obtain similar results. RQMC methods can be viewed as an alternative to MC in computing statistical bounds, as e.g. in Shapiro (2003). In our tests, the bounds for the optimal values obtained with RQMC are consistently tighter than those obtained with MC.

Other variance reduction techniques, like antithetic variates, importance - and latin hypercube sampling have been used in stochastic optimization e.g. in Kouwenberg (2001), Infanger (1992), Higle (1998) and Linderoth et al. (2002). These studies show that variance reduction techniques can significantly improve the accuracy of the sample approximations over MC. It was found in Linderoth et al. (2002) that latin hypercube sampling provides tighter confidence intervals for optimal values than MC. In our tests, the best performing RQMC methods consistently outperform latin hypercube sampling.

Since we are dealing with minimization problems, a natural framework for analyzing approximations is *epi-convergence;* see Attouch (1984) or Rockafellar and Wets (1998) for introduction to epi-convergence. Epi-convergence of the objectives is a minimal property that should be satisfied by any approximation scheme for optimization problems in order to get asymptotic convergence of optimal values and solutions. Epi-convergence for sample approximations of stochastic programs have been proved in Artstein and Wets (1995) for MC, and in Pennanen and Koivu (2003) for QMC. In MC $\{P^\nu\}_{\nu=1}^\infty$ is a sequence of empirical measures, whereas in QMC it is a weakly convergent non-random sequence. In this paper we will show that the epi-convergence result derived in Pennanen and Koivu (2003) for QMC also applies to RQMC methods.

The rest of this paper is organized as follows. Section 2 gives a brief review of the epi-convergence results that will be utilized in this paper. Section 3 reviews the used randomization technique for QMC point sets. It is shown in Section 3 that RQMC methods produce weakly convergent probability measures, thus allowing us to utilize the epi-convergence results derived in Pennanen and Koivu (2003). In Section 4 we use RQMC methods to construct epi-convergent sample approximations of stochastic programs in various test problems, and compare the behaviour of optimal values numerically with MC.

2 Epi-convergence of sample approximations

Epi-convergence results for sample approximations of stochastic optimization problems have been given in Artstein and Wets (1995) for MC, and in Pennanen and Koivu (2003) for QMC. In MC $\{P^\nu\}_{\nu=1}^\infty$ is a sequence of empirical measures, whereas in QMC it is a weakly convergent non-random sequence, that is

$$E^{P^\nu}\varphi \to E^P\varphi, \tag{1}$$

for all bounded and continuous functions φ; see Billingsley (1999). Epi-convergence has many important implications in studying approximations of minimization problems; see e.g. Rockafellar and Wets (1998). The following is one of them; see (Attouch, 1984, Section 2.2).

Theorem 1 *If a sequence of functions F^ν epi-converges to F, then*

$$\limsup_{\nu\to\infty}\ \inf F^\nu \le \inf F,$$

and if there is a convergent sequence $x^k \to x$ such that $x^k \in \operatorname{argmin} F^{\nu^k}$ for some subsequence $\{\nu^k\}_{k=1}^\infty$, then $x \in \operatorname{argmin} F$ and $\inf F^{\nu^k} \to \inf F$. In particular, if there is a compact set C such that $\operatorname{argmin} F^\nu \cap C \neq \emptyset$ for all ν, then $\inf F^\nu \to \inf F$.

Recall that a function g is called *lower semicontinuous* (lsc) if for every x

$$\liminf_{y\to x} g(y) \ge g(x).$$

Theorem 2 (Artstein and Wets (1995)) *Let $\xi_1, \xi_2, \ldots$ be a sequence of i.i.d P-distributed drawings from Ξ and let*

$$P^\nu = \sum_{i=1}^{\nu} \frac{1}{\nu}\delta_{\xi_i}.$$

If

1. *$f(x,\xi) : \mathbb{R}^n \times \Xi \to (-\infty,\infty]$ is measurable on $\mathbb{R}^n \times \Xi$, and $f(\cdot,\xi)$ for ξ fixed is lsc in x,*
2. *for each $x_0 \in \mathbb{R}^n$ there exists an open set $N \ni x_0$ and an integrable function $\alpha(\xi) : \Xi \to (-\infty,\infty)$, such that for almost all $\xi \in \Xi$ the inequality $f(x,\xi) \ge \alpha(\xi)$ holds for all $x \in N$,*

then the functions $E^{P^\nu} f$ almost surely epi-converge to $E^P f$.

The following is a simplified version of the epi-convergence result in Pennanen and Koivu (2003), which is sufficient in the applications of this paper.

Theorem 3 (Pennanen and Koivu (2003)) *Let $P^\nu \to P^0$ and assume that f is lsc. If*

1. *for each $x \in \mathbb{R}^n$, there is an open set $N \ni x$ such that f is bounded from below on $N \times \Xi$,*
2. *for each $x \in \operatorname{dom} E^{P^0} f$, $f(x, \cdot)$ is P^0-a.s. continuous and bounded,*

then the functions $E^{P^\nu} f$ both pointwise and epi-converge to $E^P f$.

Note that the conditions of Theorem 3 imply the conditions of Theorem 2.

3 Randomized quasi-monte carlo and weak convergence

A discrete approximation P^ν of P is usually generated as follows: In the scalar case, approximate the uniform distribution on $[0, 1]$ and transform each point with the inverse of the distribution function of the desired distribution. This is known as the *method of inversion.* The same idea works whenever $P = QG^{-1}$, where Q is the multivariate uniform distribution and G is Q-a.s. continuous, in other words, whenever

$$\xi = G(u),$$

where u is uniformly distributed in the unit cube $[0, 1]^d$, and $G : [0, 1]^d \to \Xi$ is almost everywhere continuous. This is based on the following very useful result from Billingsley (1999) where U is any metric space with Borel algebra $\mathcal{B}$.

Theorem 4 (Billingsley) *Let $G : (U, \mathcal{B}) \to (\Xi, \Sigma)$ be a measurable function and Q a probability distribution on $(U, \mathcal{B})$. Then $QG^{-1}(A) := Q(G^{-1}A)$ defines a probability measure on (Ξ, Σ), and if G is Q-a.s. continuous, then*

$$Q^\nu \to Q \quad \Longrightarrow \quad Q^\nu G^{-1} \to QG^{-1}.$$

Given a Q-a.s. continuous G and a discrete approximation $Q^\nu = \sum_{i=1}^{\nu} p_i^\nu \delta_{u_i^\nu}$ of Q, Theorem 4 says that the discrete measures

$$P^\nu := Q^\nu G^{-1} = \sum_{i=1}^{\nu} p_i^\nu \delta_{G(u_i^\nu)}$$

converge weakly to $P = QG^{-1}$ whenever $Q^\nu \to Q$. It is then natural to try to choose discrete approximations Q^ν which are as close as possible to the uniform distribution Q. Quasi-monte carlo methods are designed to do exactly this; see the books of Niederreiter (1992) and Sloan and Joe (1994). Much of this theory has evolved around the following notion of distance from Q.

Definition 5 *The* star-discrepancy *of a point set $\mathcal{U}_\nu = \{u_1, \ldots, u_\nu\} \subset [0, 1]^d$ is defined as*

$$D^*(\mathcal{U}_\nu) = \sup_{C \in \mathcal{C}_0} |Q^\nu(C) - Q(C)|, \tag{2}$$

where

$$Q^\nu = \sum_{i=1}^{\nu} \frac{1}{\nu} \delta_{u_i},$$

and $\mathcal{C}_0$ is the set of rectangles $C \subset [0, 1]^d$ with $0 \in C$.

The following is a direct consequence of Corollary 11 in Lucchetti et al. (1994).

Proposition 6 *For each ν, let $\mathcal{U}_\nu^\nu = \{u_1^\nu, \dots, u_\nu^\nu\}$ be point sets in the unit cube. The measures*

$$Q^\nu = \sum_{i=1}^{\nu} \frac{1}{\nu} \delta_{u_i^\nu}$$

converge weakly to the uniform distribution if and only if $D^*(\mathcal{U}_\nu^\nu) \to 0$.

Thus, if we can find point sets whose star-discrepancy approaches zero as $\nu \nearrow \infty$, we obtain weakly convergent discrete approximations of the uniform distribution. If $P = QG^{-1}$, we can then use the method of inversion to get weakly convergent discretizations of P.

In the literature of numerical integration, many methods have been developed for producing infinite sequences, which satisfy the property, that $D^*(\mathcal{U}_\nu) = O(\nu^{-1}(\ln \nu)^d)$, for all ν. Such sequences are called *low-discrepancy sequences.* The main constructions of low discrepancy sequences are due to Halton (1960), Sobol' (1967), Faure (1982) and Niederreiter (1988). The last three methods fall in the general class of (t, s)-sequences; see Niederreiter (1992). If it is not required that ν points of a $(\nu + 1)$-point quadrature are the points of the ν-point quadrature, it is possible to obtain more accurate quadratures called *low discrepancy point sets*, which satisfy $D^*(\mathcal{U}_\nu) = O(\nu^{-1}(\ln \nu)^{d-1})$. Examples of low discrepancy point sets include Hammersley point sets (Hammersley, 1960), which are easily obtained from the Halton sequence and so called (t, m, s)-nets, which are obtained by using certain parts of the points in (t, s)-sequences; see (Niederreiter, 1992, Chapter 4). Another general family of methods for generating point sets with low discrepancy are lattice rules, which are designed to take advantage of additional regularity properties of integrands; see for example Niederreiter (1992), Sloan and Joe (1994) and L'Ecuyer and Lemieux (2000).

To enable practical error estimation for QMC methods a number of randomization techniques have been proposed in the literature; see L'Ecuyer and Lemieux (2002) for an excellent survey. An easy way of randomizing any QMC point set without destroying its regular structure, suggested by Cranley and Patterson (1976), is to shift it randomly, modulo 1, with respect to all of the coordinates.

Let $\mathcal{U}_\nu = \{u_1, \dots, u_\nu\} \subset [0, 1)^d$ be a low discrepancy point set in a d-dimensional unit hypercube. Generate a point u uniformly distributed in $[0, 1)^d$ and replace every u_i in $\mathcal{U}_\nu$ with $\tilde{u}_i = (u_i + u)$ mod 1, where $i = 1, \dots, \nu$. Now $\tilde{\mathcal{U}_\nu} = \{\tilde{u}_1, \dots, \tilde{u}_\nu\}$ is a randomized point set used to approximate $[0, 1)^d$ uniform distribution. This can be repeated m times, independently, with the same $\mathcal{U}_\nu$. We thus obtain m i.i.d copies of the random variable $E^{P^\nu}\varphi$, which we denote by $E^{P_1^\nu}\varphi, \dots, E^{P_m^\nu}\varphi$. Let $\hat{\sigma}^2 = \sum_{j=1}^m (E^{P_j^\nu}\varphi - \hat{\mu})^2/(m-1)$, where $\hat{\mu} = (E^{P_1^\nu}\varphi + \ldots + E^{P_m^\nu}\varphi)/m$.

Proposition 7 (L'Ecuyer and Lemieux (2000))

$$E[E^{P_j^\nu}\varphi] = E^P\varphi \quad \textit{and} \quad E[\hat{\sigma}^2] = Var[E^{P_j^\nu}\varphi].$$

Hence, $E^{P_j^\nu}\varphi$ is an unbiased estimator of $E^P f$ and $\hat{\sigma}^2$ is an unbiased estimator of its variance. Proposition 7 holds for an arbitrary point set $\mathcal{U}_\nu$; see (L'Ecuyer and Lemieux, 2002, Tuffin, 1996).

In direct numerical integration, Monte Carlo methods achieve a convergence rate of $\nu^{-\frac{1}{2}}$; more precisely, in Monte Carlo, the standard deviation of the integration error is $Std(\varphi)\nu^{-\frac{1}{2}}$, where $Std(\varphi)$ is the standard deviation of φ. The following estimates the convergence speed for the variance of a randomized QMC estimator obtained from a low discrepancy sequence.

Theorem 8 (Tuffin (1996)) *For any low discrepancy sequence $\mathcal{U}_\nu \subset [0,1)^d$ and almost everywhere continuous and bounded function φ over $[0,1)^d$, we have*

$$Var\left(\frac{1}{\nu}\sum_{i=1}^{\nu}\varphi(\tilde{u_i})\right) = O(\nu^{-2}(\ln\nu)^{2d}).$$

In Monte Carlo, the convergence speed is independent of the dimension of the space, whereas the above convergence speed depends on the dimension, so that the actual error estimates obtained in practice with RQMC may be greater than $Std(\varphi)\nu^{-\frac{1}{2}}$. In many practical applications, however, RQMC methods considerably improve the accuracy over MC. One explanation offered for the success of QMC and RQMC methods on high dimensional problems is that the integrands may have effective dimensions much smaller than d. Effective dimension is roughly the number of important dimensions of the problem, which account for most of the variability of the estimator; see Caflisch et al. (1997) and Wang and Fang (2002) for details. Asymptotically the variance reduction factor obtained with RQMC over MC is proportional to ν. The same effect can be observed in the test problems of Section 4, for sample variances of optimal values already with moderate values of ν.

It is well known, that for MC

$$\inf_{x\in\mathbb{R}^n} E\left[E^{P^\nu} f(x)\right] \geq E\left[\inf_{x\in\mathbb{R}^n} E^{P^\nu} f(x)\right],$$

i.e. $v^* \geq E[\bar{v}^*]$ where v^* denotes the optimal value of the true problem (SP). That is, $\bar{v}^*$ is a biased estimator of v^*. This property also holds for RQMC methods. The value $\bar{v}^*$ is called a valid statistical lower bound of the true optimal value v^* if $v^* \geq E[\bar{v}^*]$ and $\bar{v}^*$ epi-converges to v^* as $\nu \to \infty$; see e.g. Shapiro (2003).

For obtaining epi-converge of the sample approximations of stochastic programs generated via RQMC methods we need to show that RQMC methods generate weakly convergent probability measures.

Lemma 9 *Let $\mathcal{U}_\nu$ and $\tilde{\mathcal{U}_\nu}$ be low discrepancy and randomized low discrepancy point sets, respectively. Discrepancy of a randomized low discrepancy point set $D(\tilde{\mathcal{U}_\nu})$ satisfies*

$$D(\tilde{\mathcal{U}_\nu}) \leq 2^{2d} D^*(\mathcal{U}_\nu).$$

If $D^*(U_\nu) \to 0$, the measures

$$Q^\nu = \sum_{i=1}^{\nu} \frac{1}{\nu}\delta_{\tilde{u}_i^\nu}$$

converge weakly to the uniform distribution.

Proof. From Niederreiter (1992) we get

$$D^*(\mathcal{U}_\nu) \leq D(\mathcal{U}_\nu) \leq 2^d D^*(\mathcal{U}_\nu),$$

where $D(\mathcal{U}_\nu)$ is a discrepancy measure. Tuffin (1996) showed that

$$D(\tilde{\mathcal{U}_\nu}) \leq 2^d D(\mathcal{U}_\nu),$$

which yields

$$D(\tilde{\mathcal{U}_\nu}) \le 2^{2d} D^*(\mathcal{U}_\nu).$$

The weak convergence of the probability measures $Q^\nu = \sum_{i=1}^{\nu} \frac{1}{\nu}\delta_{\tilde{u}_i^\nu}$ follows from Proposition 6 by noting that $D^*(U_\nu) \to 0$. □

Hence, we can use the results of Theorem 3 for obtaining epi-convergence of $E^{P^\nu} f$ to $E^P f$. In sample approximations of stochastic programs a natural goal is to try to generate the samples so that the bias $v^* - E[\bar{v}^*]$ and the sample variance of the optimal values are as small as possible. In the next Section we use RQMC methods as variance reduction techniques alone and in combination with other variance reduction techniques to improve the accuracy of sample approximations with respect to MC in various test problems.

4 Numerical tests

In the numerical tests we compare MC with variance reduction techniques: Antithetic Variates (AV), Latin Hypercube sampling (LH), randomized Lattice Rules (LR), Sobol (SOB), Faure (FAU), Hammersley (HAM), Niederreiter (NIE) and Halton (HAL) point sets in discretization of five portfolio optimization problems. We will also test the efficiency of the best performing RQMC methods in combination with AV, namely Sobol sequence (SOB+AV) and lattice rules (LR+AV). For the MC method and randomization of the QMC point sets we use the Mersenne Twister generator (MT19937) by Matsumoto and Nishimura (1998). The LIBSEQ[1] library based on Friedel and Keller (2002) is used for Latin Hypercube sampling. Rank-1 lattice rules are used to generate the lattice point sets; see e.g. L'Ecuyer and Lemieux (2000) [2]. Our implementation of the Sobol sequence is based on the implementation in Press et al. (1992). For Niederreiter sequence the routine in GSL (Gnu Scientific Library) is used. Routines by Fox (1986) are used for Faure and Halton sequences and the Hammersley point sets are easily obtained from the Halton sequence; see Hammersley (1960).

We consider one-stage problems with $\nu = 2^i$ scenarios, where $i = 5, \ldots, 14$. For every i we generate 250 independent discretizations, solve the resulting problems and record the obtained optimum value and other relevant statistics. The same procedure is repeated for each test problem, except in Section 4.2.1, where the random variable is one-dimensional and $i = 5, \ldots, 9$.

The test problems are divided into two categories. In Section 4.1 we consider problems without implicit constraints, i.e. dom $E^P f$ is known and does not depend on P. In Section 4.2 we consider problems with implicit constraints, i.e. dom $E^P f$ may not be known and may depend on P.

4.1 Problems without implicit constraints

4.1.1 Mean-variance portfolio optimization

We start the numerical tests with a model which can be solved exactly. Of course, sample approximations are unnecessary in such cases but here we get to compare the approximate solutions with

[1] `www.multires.caltech.edu/software/libseq`

[2] The parameters required by the method were provided by Professor L'Ecuyer.

the exact one. Consider the mean-variance model

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{minimize} \quad & E^{P^0}(r\cdot x-\bar r\cdot x)^2 \\ subject\ \ to \quad & \bar r\cdot x\ge w, \\ & \sum_{i=1}^n x_i\le 1, \\ & x\in C, \end{aligned} \tag{MP}$$

where $x=(x_1,\dots,x_n)$ is a portfolio of assets, $r=(r_1,\dots,r_n)$ is the vector of returns, $r\cdot x=\sum_{j=1}^n r_ix_i$ is the terminal wealth, w is the required level of expected wealth and C is the set of feasible portfolios. The components of the return vector r are random variables with joint distribution P^0 and expectation $\bar r$. As is well-known, the expectation in (MP) can be computed explicitly as

$$E^{P^0}(r\cdot x-\bar r\cdot x)^2=E^{P^0}[(r-\bar r)\cdot x]^2=E^{P^0}[x\cdot(r-\bar r)(r-\bar r)^Tx]=x\cdot Vx,$$

where $V=E^{P^0}[(r-\bar r)(r-\bar r)^T]$ is the variance matrix of r. If V and $\bar r$ are known, (MP) can then be solved without discretization with standard solvers yielding the optimal value and optimal solution.

To test the performance of the proposed variance reduction techniques, we approximate problem (MP) by the discretizations

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{minimize} \quad & \sum_{i=1}^{\nu} p_i^\nu(r_i^\nu\cdot x-\bar r\cdot x)^2 \\ subject\ \ to \quad & \bar r\cdot x\ge w, \\ & \sum_{i=1}^n x_i\le 1, \\ & x\in C. \end{aligned} \tag{MP^ν}$$

Under mild conditions, convergence of optimal values and solutions can be guaranteed. The proof of the following Proposition can be found in Pennanen and Koivu (2003).

Proposition 10 (Pennanen and Koivu (2003)) *Assume that* $\operatorname{supp} P^0$ *is bounded,* C *is closed, and that the measures*

$$P^\nu=\sum_{i=1}^{\nu} p_i^\nu\delta_{r_i^\nu}$$

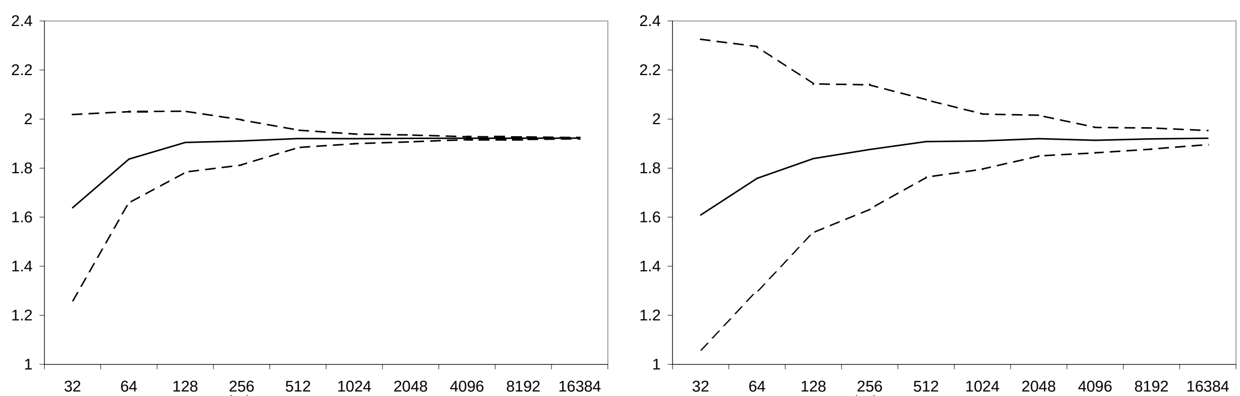
converge weakly to P^0 *and satisfy* $\operatorname{supp} P^\nu\subset\operatorname{supp} P^0$*. If the feasible set is bounded, then the optimal values of* (MP^ν) *converge to that of* (MP) *and the cluster points of the solutions of* (MP^ν) *are solutions of* (MP)*.*

In our test, the number of assets $n=10$ and

$$r=\bar r+12L(u-\frac{1}{2}e),$$

where u is uniformly distributed in the 10-dimensional unit cube, L is a 10×10 matrix and e is a vector of ones. Then supp P^0 is bounded, r has mean $\bar{r}$ and variance $V = LL^T$. We chose $C = \mathbb{R}^n_+$, which means that "short selling" is prohibited. With our choices of $\bar{r}$ and V, the optimal value in the original problem (MP) is 1.9221.

The numerical test results are displayed in Table 1, where $\hat{\mu}$ and $\hat{\sigma}$ denote the sample mean and standard deviation computed from 250 optimal values of (MP^ν) for different values of ν. The value $vr = \hat{\sigma}^2_{MC}/\hat{\sigma}^2_q$, denotes the variance reduction factors for optimal values obtained with sampling method q with respect to the variance of MC, for all the considered methods and reported values of ν. The best performing methods are LR and Sobol, Halton and Niederreiter sequences, with variance reduction factors increasing with ν. These methods clearly outperform MC, AV and LH sampling. The results with AV are presented to point out the fact, that the use of AV doubles the variance with respect to MC because the objective function is quadratic and it is well known, that AV reduces the variance compared to MC only when the integrand is a monotonically increasing function of the random variables; see (Bratley et al., 1987). Figure 1 shows the sample mean and 90% confidence intervals for the optimal values obtained with LR and MC. Lattice rules produce much tighter confidence intervals and reduces the sample bias for the optimal value, compared to MC.



(a) Lattice rule. (b) Monte carlo.

Figure 1: Mean and 90% confidence interval for the markowitz problem.

Table 1: Statistics for MP^{ν} as a function of ν.

ν		MC	AV	LH	LR	SOB	FAU	HAM	NIE	HAL
32	$\hat{\mu}$	1.609	1.413	1.668	1.639	1.662	1.461	1.567	1.708	1.704
	$\hat{\sigma}$	5.00E-1	6.99E-1	4.06E-1	3.68E-1	3.83E-1	5.76E-1	4.92E-1	3.78E-1	3.31E-1
	vr	**1.0**	**0.5**	**1.5**	**1.8**	**1.7**	**0.8**	**1.0**	**1.7**	**2.3**
64	$\hat{\mu}$	1.758	1.689	1.752	1.837	1.855	1.742	1.796	1.840	1.818
	$\hat{\sigma}$	3.38E-1	4.61E-1	2.98E-1	1.48E-1	2.05E-1	2.56E-1	2.20E-1	1.82E-1	1.72E-1
	vr	**1.0**	**0.5**	**1.3**	**5.2**	**2.7**	**1.7**	**2.4**	**3.4**	**3.9**
128	$\hat{\mu}$	1.839	1.803	1.846	1.905	1.875	1.888	1.890	1.883	1.889
	$\hat{\sigma}$	2.08E-1	2.93E-1	1.74E-1	7.73E-2	1.20E-1	1.06E-1	1.22E-1	1.16E-1	9.43E-2
	vr	**1.0**	**0.5**	**1.4**	**7.3**	**3.0**	**3.9**	**2.9**	**3.2**	**4.9**
256	$\hat{\mu}$	1.876	1.820	1.887	1.911	1.906	1.904	1.909	1.916	1.913
	$\hat{\sigma}$	1.53E-1	2.22E-1	1.15E-1	5.78E-2	6.93E-2	5.78E-2	6.33E-2	6.19E-2	5.52E-2
	vr	**1.0**	**0.5**	**1.8**	**7.1**	**4.9**	**7.1**	**5.9**	**6.1**	**7.7**
512	$\hat{\mu}$	1.908	1.877	1.899	1.921	1.920	1.909	1.916	1.914	1.914
	$\hat{\sigma}$	1.04E-1	1.38E-1	8.48E-2	2.15E-2	3.45E-2	4.48E-2	3.75E-2	3.53E-2	2.84E-2
	vr	**1.0**	**0.6**	**1.5**	**23.5**	**9.1**	**5.4**	**7.8**	**8.7**	**13.5**
1024	$\hat{\mu}$	1.911	1.902	1.917	1.920	1.920	1.920	1.921	1.923	1.922
	$\hat{\sigma}$	7.09E-2	1.02E-1	5.87E-2	1.24E-2	1.81E-2	2.10E-2	1.97E-2	1.89E-2	1.67E-2
	vr	**1.0**	**0.5**	**1.5**	**32.7**	**15.3**	**11.4**	**12.9**	**14.0**	**18.1**
2048	$\hat{\mu}$	1.920	1.906	1.920	1.922	1.921	1.921	1.923	1.922	1.922
	$\hat{\sigma}$	4.98E-2	7.02E-2	4.15E-2	8.17E-3	8.87E-3	1.32E-2	1.01E-2	1.02E-2	9.34E-3
	vr	**1.0**	**0.5**	**1.4**	**37.1**	**31.5**	**14.2**	**24.1**	**23.8**	**28.4**
4096	$\hat{\mu}$	1.914	1.917	1.920	1.922	1.922	1.922	1.922	1.922	1.922
	$\hat{\sigma}$	3.29E-2	5.25E-2	2.89E-2	3.65E-3	4.69E-3	6.97E-3	5.29E-3	6.43E-3	5.90E-3
	vr	**1.0**	**0.4**	**1.3**	**80.9**	**49.0**	**22.2**	**38.6**	**26.1**	**31.1**
8192	$\hat{\mu}$	1.919	1.921	1.923	1.922	1.922	1.922	1.922	1.922	1.922
	$\hat{\sigma}$	2.61E-2	3.57E-2	2.02E-2	3.33E-3	3.32E-3	3.55E-3	2.83E-3	3.59E-3	2.94E-3
	vr	**1.0**	**0.5**	**1.7**	**61.4**	**61.7**	**54.1**	**85.3**	**52.8**	**78.8**
16384	$\hat{\mu}$	1.922	1.920	1.923	1.922	1.922	1.922	1.922	1.922	1.922
	$\hat{\sigma}$	1.79E-2	2.46E-2	1.42E-2	1.49E-3	1.37E-3	1.84E-3	1.53E-3	1.18E-3	1.78E-3
	vr	**1.0**	**0.5**	**1.6**	**145**	**170**	**94.5**	**137**	**229**	**101**

4.1.2 Utility maximization

Consider the problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{maximize} \quad & E^{P^0} u\left(r\cdot x\right) \\ subject\ \ to \quad & \sum_{i=1}^{n} x_i \leq 1, \\ & x \in C. \end{aligned} \tag{UP}$$

Here x and C are as in the previous example, u measures the utility from terminal wealth, and the components of the return vector r are nonnegative random variables with joint distribution P^0.

In general, (UP) cannot be solved analytically, so we consider the discretizations

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{maximize} \quad & \sum_{i=1}^{\nu} p_i^\nu u(r_i^\nu \cdot x) \\ subject\ \ to \quad & \sum_{i=1}^{n} x_i \leq 1, \\ & x \in C. \end{aligned} \tag{UP^ν}$$

The same type of problem was analyzed in Pennanen and Koivu (2003), so we can use their Proposition to show the epi-convergence of (UP^ν) to (UP).

Proposition 11 (Pennanen and Koivu (2003)) *Assume* $\operatorname{supp} P^0 \subset \mathbb{R}^n_+$*, u is continuous and bounded on $\mathbb{R}_+$, C is closed and contained in $\mathbb{R}^n_+$ (short selling is not allowed) and that the measures*

$$P^\nu = \sum_{i=1}^{\nu} p_i^\nu \delta_{r_i^\nu}$$

converge weakly to P^0 and satisfy $\operatorname{supp} P^\nu \subset \mathbb{R}^n_+$*. Then the optimal values of (UP^ν) converge to that of (UP) and the cluster points of the solutions of (UP^ν) are solutions of (UP).*

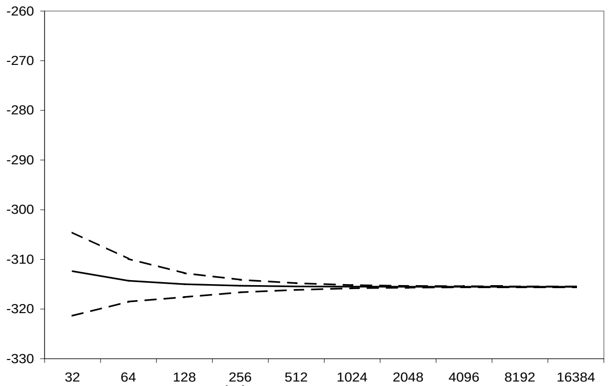
In the test, the number of assets $n = 10$, r is log-normally distributed, $u(w) = -\exp(-w)$ and $C = \mathbb{R}^n_+$. Table 2 summarizes the test results. AV reduces the bias and variance of the optimal values significantly compared to MC. Among the RQMC methods LR perform the best, with all the other quadratures, except Faure sequence, performing almost as well. Since the use of AV reduced the variance of optimal values considerably, we tested them in combination with LR and Sobol sequence, see Table 3. The combination of these methods produce the most significant variance reduction factors compared to MC. Figure 2 displays the sample mean and 90% confidence interval for the optimal values obtained with LR and MC. Again the variance reduction factors with RQMC methods increase almost linearly with ν.

Table 2: Statistics UP^{ν} as a function of ν.

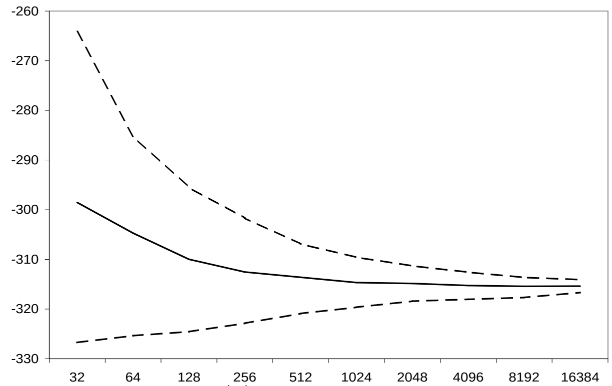
ν		MC	AV	LH	LR	SOB	FAU	HAM	HAL	NIE
32	$\hat{\mu}$	-298.543	-312.304	-312.014	-312.372	-310.990	-306.122	-308.435	-306.970	-310.261
	$\hat{\sigma}$	1.94E+1	3.84E+0	7.10E+0	5.18E+0	6.90E+0	1.20E+1	5.89E+0	5.82E+0	5.85E+0
	vr	**1.0**	**25.6**	**7.5**	**14.0**	**7.9**	**2.6**	**10.8**	**11.1**	**11.0**
64	$\hat{\mu}$	-304.718	-313.774	-313.855	-314.307	-313.418	-311.252	-312.172	-311.824	-313.258
	$\hat{\sigma}$	1.26E+1	2.67E+0	3.55E+0	2.77E+0	3.78E+0	7.26E+0	3.56E+0	3.33E+0	3.13E+0
	vr	**1.0**	**22.3**	**12.6**	**20.8**	**11.2**	**3.0**	**12.6**	**14.4**	**16.3**
128	$\hat{\mu}$	-309.980	-314.782	-314.865	-314.945	-314.853	-314.599	-314.174	-313.945	-314.681
	$\hat{\sigma}$	9.07E+0	1.72E+0	1.92E+0	1.38E+0	2.02E+0	2.88E+0	1.81E+0	1.69E+0	1.54E+0
	vr	**1.0**	**27.7**	**22.4**	**43.1**	**20.2**	**9.9**	**25.2**	**29.0**	**34.8**
256	$\hat{\mu}$	-312.546	-315.173	-315.171	-315.303	-315.201	-315.295	-315.136	-314.915	-315.236
	$\hat{\sigma}$	6.78E+0	1.09E+0	1.04E+0	7.66E-1	9.55E-1	1.86E+0	9.40E-1	8.99E-1	8.28E-1
	vr	**1.0**	**38.9**	**42.4**	**78.4**	**50.4**	**13.2**	**52.0**	**56.9**	**67.1**
512	$\hat{\mu}$	-313.600	-315.151	-315.368	-315.448	-315.475	-315.289	-315.350	-315.357	-315.389
	$\hat{\sigma}$	4.40E+0	9.27E-1	5.33E-1	4.17E-1	4.77E-1	1.23E+0	4.78E-1	4.79E-1	4.07E-1
	vr	**1.0**	**22.5**	**68.1**	**111**	**84.9**	**12.8**	**84.6**	**84.2**	**116**
1024	$\hat{\mu}$	-314.658	-315.443	-315.414	-315.489	-315.502	-315.411	-315.440	-315.471	-315.457
	$\hat{\sigma}$	3.07E+0	6.18E-1	3.06E-1	2.02E-1	2.60E-1	5.66E-1	2.68E-1	2.44E-1	2.11E-1
	vr	**1.0**	**24.7**	**100**	**232**	**139**	**29.4**	**131**	**158**	**211**
2048	$\hat{\mu}$	-314.858	-315.440	-315.465	-315.505	-315.512	-315.473	-315.492	-315.486	-315.498
	$\hat{\sigma}$	2.15E+0	4.58E-1	1.74E-1	1.07E-1	1.37E-1	3.07E-1	1.32E-1	1.40E-1	1.10E-1
	vr	**1.0**	**22.1**	**153**	**404**	**245**	**49.0**	**264**	**235**	**381**
4096	$\hat{\mu}$	-315.270	-315.475	-315.481	-315.506	-315.506	-315.482	-315.509	-315.498	-315.496
	$\hat{\sigma}$	1.70E+0	2.94E-1	1.16E-1	5.89E-2	7.61E-2	1.57E-1	7.60E-2	7.01E-2	6.13E-2
	vr	**1.0**	**33.3**	**216**	**832**	**498**	**118**	**500**	**587**	**767**
8192	$\hat{\mu}$	-315.421	-315.483	-315.495	-315.505	-315.502	-315.493	-315.508	-315.503	-315.504
	$\hat{\sigma}$	1.23E+0	2.02E-1	6.62E-2	3.16E-2	3.78E-2	9.33E-2	4.05E-2	4.25E-2	3.20E-2
	vr	**1.0**	**37.5**	**348**	**1527**	**1062**	**175**	**926**	**844**	**1482**
16384	$\hat{\mu}$	-315.381	-315.495	-315.495	-315.504	-315.505	-315.511	-315.503	-315.506	-315.504
	$\hat{\sigma}$	8.07E-1	1.58E-1	4.89E-2	2.00E-2	1.91E-2	4.66E-2	2.08E-2	2.11E-2	1.90E-2
	vr	**1.0**	**25.9**	**272**	**1630**	**1791**	**299**	**1510**	**1462**	**1804**

Table 3: Statistics for UP^{ν} as a function of ν, Lattice rule and Sobol with AV.

ν		LR+AV	SOB+AV	ν		LR+AV	SOB+AV
32	$\hat{\mu}$	-313.078	-312.744	1024	$\hat{\mu}$	-315.485	-315.488
	$\hat{\sigma}$	1.61E+0	2.26E+0		$\hat{\sigma}$	1.13E-1	1.33E-1
	vr	**145**	**74**		vr	**737**	**528**
64	$\hat{\mu}$	-314.278	-314.295	2048	$\hat{\mu}$	-315.498	-315.496
	$\hat{\sigma}$	1.15E+0	1.31E+0		$\hat{\sigma}$	5.94E-2	7.43E-2
	vr	**121**	**93**		vr	**1312**	**837**
128	$\hat{\mu}$	-314.962	-315.040	4096	$\hat{\mu}$	-315.499	-315.503
	$\hat{\sigma}$	5.79E-1	6.35E-1		$\hat{\sigma}$	3.91E-2	4.71E-2
	vr	**246**	**204**		vr	**1886**	**1304**
256	$\hat{\mu}$	-315.324	-315.317	8192	$\hat{\mu}$	-315.505	-315.507
	$\hat{\sigma}$	3.16E-1	3.75E-1		$\hat{\sigma}$	2.75E-2	2.45E-2
	vr	**460**	**327**		vr	**2007**	**2532**
512	$\hat{\mu}$	-315.415	-315.429	16384	$\hat{\mu}$	-315.504	-315.505
	$\hat{\sigma}$	2.37E-1	2.16E-1		$\hat{\sigma}$	1.41E-2	1.50E-2
	vr	**343**	**416**		vr	**3295**	**2889**



(a) Lattice rule.



(b) Monte carlo.

Figure 2: Mean and 90% confidence interval for the utility maximization problem.

4.1.3 Hedging with contingent claims

Assume that a company's operating revenue at time $t = 0, \ldots, T$ can be expressed as a function $\pi_t(\xi)$, where $\xi = (\xi_0, \ldots, \xi_T)$ is a stochastic process with joint distribution P^0. The company wishes to hedge its operating revenue against unfavorable outcomes of ξ using contingent claims with pay-outs $F_t(\xi)$. Let $\theta_+ = (\theta_1, \ldots, \theta_J)$ and $\theta_- = (\theta_1, \ldots, \theta_J)$ denote the amounts of contingent claims bought and sold with prices P_a and P_b, respectively, at time $t = 0$. The company faces the hedging problem

$$\begin{aligned} \underset{\theta_+,\theta_-}{maximize} \quad & E^{P^0}\left[u\left(\pi_0(\xi_0) - tc_a \cdot \theta_+ - tc_b \cdot \theta_-\right) + \sum_{t=1}^{T} u\left(\pi_t(\xi) + F_t(\xi) \cdot (\theta_+ - \theta_-)\right)\right] \qquad (HP) \\ subject\ \ to \quad & P_a \cdot \theta_+ - P_b \cdot \theta_- \leq \pi_0(\xi_0) \\ & \qquad \theta_+, \theta_- \geq 0, \end{aligned}$$

where u is a utility function, $\pi_0(\xi_0)$ is fixed and tc_a and tc_b denote the transaction costs of bought and sold assets, respectively. Since (HP) is impossible to solve analytically we consider the discretizations

$$\begin{aligned} \underset{\theta_+,\theta_-}{maximize} \quad & \sum_{i=1}^{\nu} p_i^{\nu}\left[u\left(\pi_0(\xi_0) - tc_a \cdot \theta_+ - tc_b \cdot \theta_-\right) + \sum_{t=1}^{T} u\left(\pi_t(\xi_i^{\nu}) + F_t(\xi_i^{\nu}) \cdot (\theta_+ - \theta_-)\right)\right] \quad (HP^{\nu}) \\ subject\ \ to \quad & P_a \cdot \theta_+ - P_b \cdot \theta_- \leq \pi_0(\xi_0) \\ & \qquad \theta_+, \theta_- \geq 0. \end{aligned}$$

Proposition 12 *Assume that u is continuous and concave, the first moments of the random variables $\pi_t(\xi)$ and $F_t(\xi)$ exist and*

$$P^\nu = \sum_{i=1}^{\nu} p_i^\nu \delta_{(\xi_{t,i}^\nu)_{t=1}^T}$$

is a sequence of empirical measures. Then with probability one the optimal values of (HP^ν) converge to that of (HP) and the cluster points of the solutions of (HP^ν) are solutions of (HP).

Proof. This can be written as (SP) with $x = (\theta_+, \theta_-)$ and

$$f(x,\xi) = -u\left(\pi_0(\xi_0) - tc_a \cdot \theta_+ - tc_b \cdot \theta_-\right) - \sum_{t=1}^{T} u\left(\pi_t(\xi) + F_t(\xi) \cdot (\theta_+ - \theta_-)\right) + \delta_{C'}(\theta_+, \theta_-),$$

where $C' = \left\{(\theta_+, \theta_-) \in \mathbb{R}_+^n \mid P_a \cdot \theta_+ - P_b \cdot \theta_- \leq \pi_0(\xi_0)\right\}$. By Theorem 1 it suffices to verify the conditions of Theorem 2. Since u is continuous and $\pi_t(\xi)$ and $F_t(\xi)$ are measurable f is measurable and lsc in x. To verify condition 2 let (x^0, ξ^0) be such that $f(x^0, \xi^0) < \infty$. By convexity of $-u$ we have

$$f(x,\xi) \geq f(x^0,\xi^0) + \gamma_0^0 \left(tc_a \cdot (\theta_+^0 - \theta_+) + tc_b \cdot (\theta_-^0 - \theta_-)\right) + \sum_{t=1}^{T} \gamma_t^0 (\pi_t(\xi) + F_t(\xi) \cdot (\theta_+ - \theta_-) - \\ \pi(\xi^0) - F_t(\xi^0) \cdot (\theta_+^0 - \theta_-^0)),$$

where γ_t^0 denote subgradients of $-u$. Using the Cauchy-Schwarz inequality we get that for any bounded $N \ni x^0$

$$f(x,\xi) \geq \psi^0 - \gamma_0^0 \left(tc_a \cdot \theta_+ + tc_b \cdot \theta_-\right) + \sum_{t=1}^{T} \gamma_t^0 (\pi_t(\xi) + F_t(\xi) \cdot (\theta_+ - \theta_-)) \geq \\ a + \sum_{t=1}^{T} \gamma_t^0 \pi_t(\xi) + b \sum_{t=1}^{T} |F_t(\xi)|, \forall x \in N,$$

where a and b are constants. Since it was assumed that the first moments of the random variables $\pi_t(\xi)$ and $F_t(\xi)$ exist condition 2 is satisfied. □

By assuming that $\pi_t(\xi)$ and $F_t(\xi)$ are almost everywhere continuous and bounded, the conditions of Theorem 3 would be satisfied and we would obtain epi-convergence for RQMC methods. However, it is interesting to study the behavior of RQMC methods in this problem numerically. In the test $u(w) = -\exp(-w)$, $T = 12$, ξ_0 is deterministic and ξ_t is a three dimensional log-normally distributed random variable, which means that the dimension of the probability space, $d = 36$. The stochastic factors affecting the company's operating revenue are the Euro-U.S. dollar (USD), Norwegian krone-USD exchange rates and the USD price of zinc. The set of contingent claims consists of zero coupon bonds and futures contracts for the underlying stochastic factors, with maturities $1, 2, \ldots, T$ months.

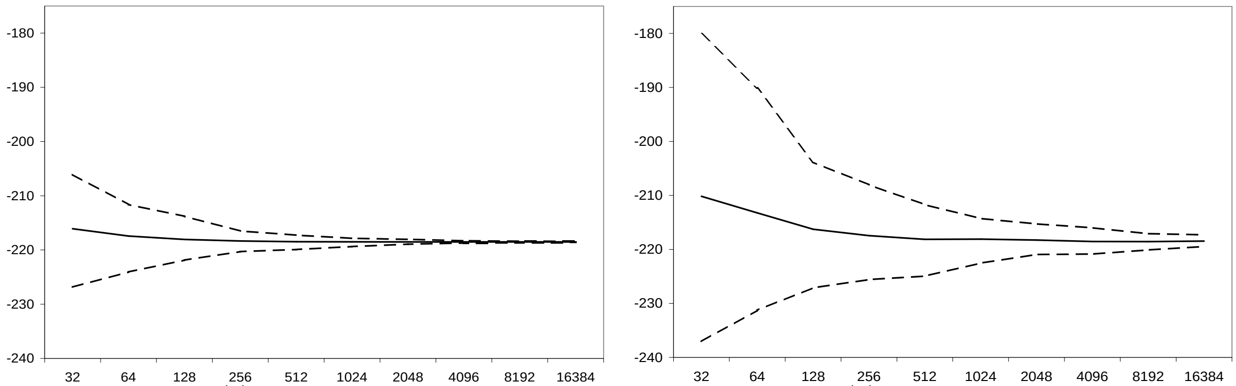
The results are displayed in Table 4. The use of AV increased the variance of optimal values compared to MC, because the profit function $\pi_t(\xi)$ is not a monotonically increasing function of the random variables, these results are not reported. The results for Niederreiter sequence are missing, because in our implementation of the sequence the maximum dimension for the probability space is 12. Again RQMC methods, except Faure sequence, clearly beat MC and LR seem to perform slightly better than the other RQMC methods. In this problem Faure sequence performs poorly and even loses to MC for low values of ν. Latin hypercube sampling substantially improves the performance over MC. Compared to LH, LR and Sobol sequence produce more accurate estimates for optimal values, for all values of ν. Figure 3 displays the sample mean and 90% confidence interval for the optimal values obtained with LR and MC, which also shows that LR clearly outperform MC. The sample means of optimal values seem to converge toward a common value with all the methods, even though we were able to proof the epi-convergence only for MC.

Table 4: Statistics for HP as a function of ν.

ν		MC	LH	LR	SOB	FAU	HAM	HAL
32	$\hat{\mu}$	-210.165	-214.596	-216.088	-215.000	-202.903	-205.388	-203.808
	$\hat{\sigma}$	1.77E+1	8.87E+0	6.69E+0	8.17E+0	2.09E+1	1.26E+1	1.31E+1
	vr	**1.0**	**4.0**	**7.0**	**4.7**	**0.7**	**2.0**	**1.8**
64	$\hat{\mu}$	-213.241	-216.771	-217.430	-217.106	-207.698	-215.237	-215.076
	$\hat{\sigma}$	1.22E+1	4.69E+0	3.87E+0	3.82E+0	1.71E+1	5.11E+0	5.04E+0
	vr	**1.0**	**6.8**	**10.0**	**10.2**	**0.5**	**5.7**	**5.9**
128	$\hat{\mu}$	-216.275	-217.842	-218.051	-217.945	-211.040	-217.492	-217.844
	$\hat{\sigma}$	7.35E+0	3.04E+0	2.38E+0	2.57E+0	1.42E+1	2.52E+0	3.10E+0
	vr	**1.0**	**5.8**	**9.6**	**8.2**	**0.3**	**8.5**	**5.6**
256	$\hat{\mu}$	-217.459	-218.298	-218.332	-218.198	-211.715	-218.202	-218.196
	$\hat{\sigma}$	5.28E+0	1.90E+0	1.19E+0	1.43E+0	1.16E+1	1.53E+0	1.47E+0
	vr	**1.0**	**7.8**	**19.7**	**13.6**	**0.2**	**11.8**	**12.9**
512	$\hat{\mu}$	-218.143	-218.461	-218.482	-218.433	-215.089	-218.372	-218.398
	$\hat{\sigma}$	4.06E+0	1.15E+0	8.14E-1	7.79E-1	7.46E+0	8.59E-1	9.04E-1
	vr	**1.0**	**12.5**	**25**	**27.2**	**0.3**	**22.4**	**20.2**
1024	$\hat{\mu}$	-218.101	-218.402	-218.496	-218.455	-218.264	-218.455	-218.423
	$\hat{\sigma}$	2.60E+0	7.59E-1	4.42E-1	5.01E-1	2.19E+0	4.79E-1	5.77E-1
	vr	**1.0**	**11.8**	**35**	**27**	**1.4**	**30**	**20**
2048	$\hat{\mu}$	-218.272	-218.484	-218.505	-218.494	-218.212	-218.490	-218.501
	$\hat{\sigma}$	1.76E+0	4.81E-1	2.79E-1	2.82E-1	1.51E+0	2.98E-1	3.06E-1
	vr	**1.0**	**13.4**	**40**	**39**	**1.4**	**35**	**33**
4096	$\hat{\mu}$	-218.535	-218.484	-218.513	-218.496	-218.419	-218.512	-218.516
	$\hat{\sigma}$	1.41E+0	3.80E-1	1.26E-1	1.59E-1	6.31E-1	1.70E-1	1.86E-1
	vr	**1.0**	**13.8**	**125**	**79**	**5**	**69**	**58**
8192	$\hat{\mu}$	-218.553	-218.504	-218.508	-218.507	-218.450	-218.507	-218.509
	$\hat{\sigma}$	9.54E-1	2.46E-1	8.34E-2	8.79E-2	5.60E-1	8.89E-2	1.01E-1
	vr	**1.0**	**15.1**	**131**	**118**	**3**	**115**	**90**
16384	$\hat{\mu}$	-218.455	-218.505	-218.514	-218.510	-218.489	-218.508	-218.511
	$\hat{\sigma}$	6.72E-1	1.70E-1	6.63E-2	5.29E-2	4.84E-1	5.26E-2	6.27E-2
	vr	**1.0**	**15.6**	**103**	**161**	**2**	**164**	**115**

4.2 Problems with implicit constraints

In the remaining examples, the feasible regions depend on the probability measure. These problems do not fit the frameworks of Lucchetti and Wets (1993), Artstein and Wets (1994), Zervos (1999) or Shapiro (2000).



(a) Lattice rule. (b) Monte carlo.

Figure 3: Mean and 90% confidence interval for the hedging problem.

4.2.1 Super-replication of contingent claims

Consider the problem

$$\begin{aligned} \underset{V,\theta}{minimize} \quad & V \\ subject\ to \quad & S_0 \cdot \theta \leq V, \\ & S \cdot \theta \geq F, \quad P^0\text{-}a.s. \\ & \theta \in C, \end{aligned} \tag{PP}$$

where V is the wealth invested in a portfolio $\theta = (\theta_1, \dots, \theta_J)$ of assets that have prices $S_0 = (S_0^1, \dots, S_0^J)$ at the beginning and $S = (S^1, \dots, S^J)$ at the end of a holding period and F is a cash-flow at the end of the holding period. S and F are random variables with joint distribution P^0. (PP) is a semi-infinite linear programming problem and, in general, impossible to solve analytically. Replacing P^0 by a discrete measure $P^\nu = \sum_{i=1}^{\nu} p_i^\nu \delta_{(S_i^\nu, F_i^\nu)}$ with $p_i^\nu > 0$, for all $i = 1, \dots \nu$ yields

$$\begin{aligned} \underset{V,\theta}{minimize} \quad & V \\ subject\ to \quad & S_0 \cdot \theta \leq V, \\ & S_i^\nu \cdot \theta \geq F_i^\nu, \quad i = 1, \dots, \nu, \\ & \theta \in C, \end{aligned} \tag{PP^ν}$$

which is an LP problem for which many solvers are available.

Proposition 13 (Pennanen and Koivu (2003)) *Assume that the points $\{(S_i^\nu, F_i^\nu)\}_{i=1}^\nu$ are all contained in* supp P^0 *and that for some $\{p_i^\nu\}_{i=1}^\nu$, $\nu = 0, 1, 2, \dots$, with $p_i^\nu > 0$, for all $i = 1, \dots \nu$, the measures*

$$P^\nu = \sum_{i=1}^{\nu} p_i^\nu \delta_{(S_i^\nu, F_i^\nu)}$$

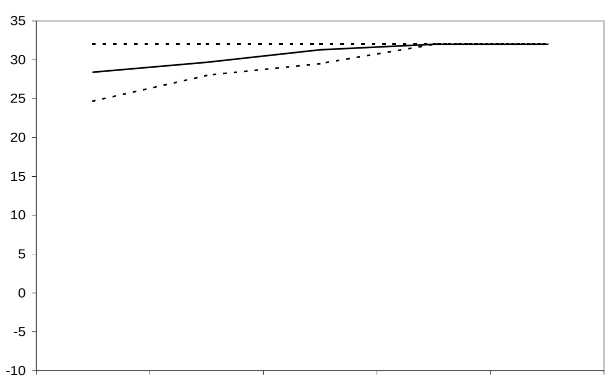
converge weakly to P^0. If the feasible set is bounded, then the optimal values of (PP^ν) converge to that of (PP) and the cluster points of the solutions of (PP^ν) are solutions of (PP).

In our test, the set of assets consists of cash, SP500 index and 28 European call and put options on the index with maturity of 17 calendar days. The value of S is fully determined by the value of the index at the maturity which is assumed to be log-normally distributed. The cash-flow F is taken to be that of a call option with the same maturity but different strike than any other call included in S.

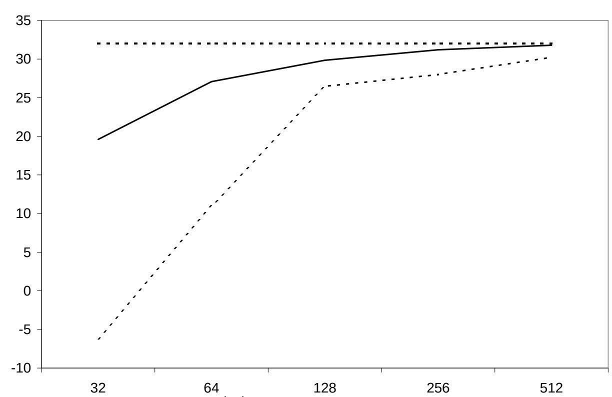
Since the random variable in this problem is one dimensional all the QMC methods produce identical discretizations. As a result we consider discretizations only with LR, AV and MC. Table 5 displays the test results. The use of AV does not improve the performance over MC. Lattice rules reduce the variance of optimal values considerably and with 256 scenarios the optimal values have converged. Figure 4 displays the average and 90% confidence interval for optimum values of (PP^{ν}) obtained with LR and MC, for each value of $\nu = 2^i, i = 5, 6, \ldots, 9$. With LR the confidence interval is much tighter and the optimal value converges faster than with MC.

Table 5: Statistics for PP as a function of ν.

ν		MC	AV	LR
32	$\hat{\mu}$	19,598	17,320	28,417
	$\hat{\sigma}$	1,71E+1	2,37E+1	2,25E+0
	vr	**1,0**	**0,5**	**57,4**
64	$\hat{\mu}$	27,072	26,739	29,682
	$\hat{\sigma}$	6,27E+0	6,87E+0	1,74E+0
	vr	**1,0**	**0,8**	**13,0**
128	$\hat{\mu}$	29,844	30,261	31,287
	$\hat{\sigma}$	3,18E+0	2,12E+0	9,05E-1
	vr	**1,0**	**2,2**	**12,3**
256	$\hat{\mu}$	31,194	31,177	32,004
	$\hat{\sigma}$	1,39E+0	1,37E+0	0,00E+0
	vr	**1,0**	**1,0**	∞
512	$\hat{\mu}$	31,786	31,841	32,004
	$\hat{\sigma}$	7,44E-1	6,00E-1	0,00E+0
	vr	**1,0**	**1,5**	∞



(a) Lattice rule.



(b) Monte Carlo.

Figure 4: Mean and 90% confidence interval for the hedging problem.

4.2.2 Utility maximization with wealth constraint

Consider the following problem

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{maximize} \quad & E^{P^0} u\,(r\cdot x) \\ subject\ \ to \quad & \sum_{i=1}^{n} x_i \le 1, \\ & x \in C \\ & r\cdot x \ge 0, \quad P^0\text{-}a.s. \end{aligned} \tag{WP}$$

This problem is a modification of the utility maximization problem of Section 4.1.2. Here $C \subset \mathbb{R}^n$, so short selling is allowed but we have added a constraint, which requires the final wealth to be almost surely non-negative. Here we are interested in studying how the short selling affects the behavior of the optimal values and solutions. The function u measures the utility from terminal wealth and the components of the return vector r are random variables with joint distribution P^0. Discretization of (WP) yields

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{maximize} \quad & \sum_{i=1}^{\nu} p_i^\nu u(r_i^\nu \cdot x) \\ subject\ \ to \quad & \sum_{i=1}^{n} x_i \le 1, \\ & x \in C, \\ & r_i^\nu \cdot x \ge 0, \quad i = 1,\dots,\nu. \end{aligned} \tag{WP^ν}$$

Proposition 14 *Assume u is continuous, nondecreasing and bounded on $\mathbb{R}_+$, C is closed and that the measures*

$$P^\nu = \sum_{i=1}^{\nu} p_i^\nu \delta_{r_i^\nu}$$

converge weakly to P^0 and satisfy $\operatorname{supp} P^\nu \subset \operatorname{supp} P^0$. *Then the optimal values of (WP^ν) converge to that of (WP) and the cluster points of the solutions of (WP^ν) are solutions of (WP).*

Proof. This fits the format of (SP) with $\Xi = \operatorname{supp} P^0$, $\xi = r$, and

$$f(x,r) = -u(r\cdot x) + \delta_{C_1}(x) + \delta_{C_2}(x,r),$$

where

$$C_1 = \left\{ x \in C \;\middle|\; \sum_{i=1}^{n} x_i \le 1 \right\}$$

and

$$C_2 = \{(x,r) \mid r\cdot x \ge 0\}.$$

We need to verify the conditions of Theorem 3. Since u is continuous, C_1 and C_2 are closed, f is lsc. Condition 1 follows from the fact that u is nondecreasing and bounded on $\mathbb{R}_+$. To verify condition 2 note that for each $x \in \operatorname{dom} E^P f$ by continuity of the inner product the requirement $r \cdot x \geq 0, P^0$-$a.s.$ is equivalent to $r \cdot x \geq 0, \forall\, r \in \operatorname{supp} P^0$. Since $\operatorname{supp} P^\nu \subset \operatorname{supp} P^0$ Condition 2 follows from the boundedness and continuity of u on $\mathbb{R}_+$. □

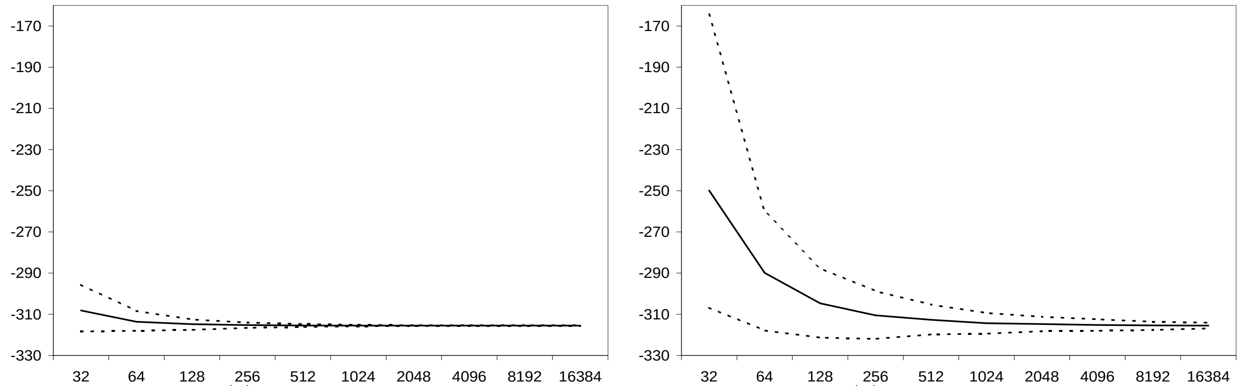
In the test $\operatorname{supp} r = \mathbb{R}^n_+$, which together with the wealth constraint implies that, $\operatorname{dom} E^P f$ is $\mathbb{R}^n_+$, so this problem differs from the utility maximization problem of Section 4.1.2 only in finite samples. The numerical test results are presented in Tables 6 and 7. The results are similar to those of Section 4.1.2. The use of antithetic variates reduces the variance considerably. When no other variance reduction technique is used, LR, Sobol and Niederreiter sequences perform the best and they reduce the variance by a factor as large as 2000. The combination of AV with LR and Sobol sequence are again the most efficient techniques; see Table 7. As expected, the sample average of optimal values converges to the same value as in the utility maximization problem of Section 4.1.2. Expected value and 90% confidence interval for the optimal values obtained with LR and MC are shown in Figure 5. In this problem, LR reduce the sample bias by a large factor and produce very thight confidence intervals for the optimal value. We characterize the infeasibility of the optimal solutions with implicit constraints by the amount of short selling in each discretized problem. The sample mean and 90% confidence interval for the maximum amount of short selling in the optimal portfolios for LR and MC are shown in Figure 6. With LR the minimum investment proportion converges towards zero much faster than with MC.

Table 6: Statistics for (WP^ν) as a function of ν.

ν		MC	AV	LH	LR	SOB	FAU	HAM	HAL	NIE
32	$\hat{\mu}$	-249.881	-305.066	-306.359	-308.130	-305.050	-283.060	-290.676	-289.656	-302.463
	$\hat{\sigma}$	4.64E+1	8.81E+0	9.47E+0	7.56E+0	9.20E+0	2.55E+1	2.05E+1	1.97E+1	1.01E+1
	vr	**1.0**	**27.7**	**24.0**	**37.6**	**25.5**	**3.3**	**5.1**	**5.6**	**21.1**
64	$\hat{\mu}$	-289.891	-312.191	-312.484	-313.656	-312.167	-306.806	-308.114	-307.586	-311.766
	$\hat{\sigma}$	1.84E+1	3.63E+0	4.08E+0	2.99E+0	4.10E+0	8.74E+0	5.80E+0	5.23E+0	3.74E+0
	vr	**1.0**	**25.7**	**20.3**	**37.8**	**20.1**	**4.4**	**10.0**	**12.3**	**24.1**
128	$\hat{\mu}$	-304.675	-314.157	-314.529	-314.818	-314.600	-314.295	-312.871	-312.464	-314.280
	$\hat{\sigma}$	1.13E+1	2.03E+0	2.05E+0	1.60E+0	2.09E+0	2.95E+0	2.50E+0	2.46E+0	1.68E+0
	vr	**1.0**	**31.1**	**30.2**	**50.0**	**29.1**	**14.6**	**20.4**	**21.1**	**45.0**
256	$\hat{\mu}$	-310.518	-314.955	-315.057	-315.253	-315.143	-315.198	-314.895	-314.597	-315.161
	$\hat{\sigma}$	7.25E+0	1.20E+0	1.08E+0	7.77E-1	9.65E-1	1.88E+0	1.03E+0	1.03E+0	8.51E-1
	vr	**1.0**	**36.7**	**44.8**	**86.9**	**56.4**	**14.8**	**49.6**	**49.2**	**72.5**
512	$\hat{\mu}$	-312.667	-315.069	-315.333	-315.435	-315.462	-315.238	-315.311	-315.278	-315.373
	$\hat{\sigma}$	4.62E+0	9.56E-1	5.44E-1	4.20E-1	4.80E-1	1.23E+0	4.93E-1	5.06E-1	4.12E-1
	vr	**1.0**	**23.4**	**72.2**	**121**	**92.6**	**14.1**	**87.7**	**83.5**	**126**
1024	$\hat{\mu}$	-314.353	-315.408	-315.401	-315.486	-315.498	-315.400	-315.425	-315.455	-315.452
	$\hat{\sigma}$	3.13E+0	6.34E-1	3.12E-1	2.02E-1	2.60E-1	5.67E-1	2.74E-1	2.53E-1	2.12E-1
	vr	**1.0**	**24.3**	**101**	**239**	**144**	**30.4**	**130**	**153**	**217**
2048	$\hat{\mu}$	-314.729	-315.424	-315.459	-315.504	-315.511	-315.470	-315.489	-315.480	-315.496
	$\hat{\sigma}$	2.18E+0	4.66E-1	1.77E-1	1.07E-1	1.38E-1	3.07E-1	1.34E-1	1.43E-1	1.11E-1
	vr	**1.0**	**21.9**	**152**	**412**	**250**	**50.4**	**266**	**233**	**389**
4096	$\hat{\mu}$	-315.208	-315.469	-315.479	-315.505	-315.505	-315.481	-315.508	-315.497	-315.502
	$\hat{\sigma}$	1.71E+0	2.96E-1	1.17E-1	5.93E-2	7.62E-2	1.57E-1	7.65E-2	7.08E-2	5.69E-2
	vr	**1.0**	**33.5**	**214**	**835**	**505**	**119**	**500**	**584**	**905**
8192	$\hat{\mu}$	-315.394	-315.480	-315.494	-315.502	-315.501	-315.492	-315.507	-315.502	-315.503
	$\hat{\sigma}$	1.23E+0	2.02E-1	6.66E-2	3.17E-2	3.80E-2	9.34E-2	4.09E-2	4.28E-2	3.24E-2
	vr	**1.0**	**37.3**	**344**	**1512**	**1057**	**175**	**912**	**833**	**1454**
16384	$\hat{\mu}$	-315.492	-315.494	-315.495	-315.506	-315.505	-315.511	-315.503	-315.506	-315.506
	$\hat{\sigma}$	8.68E-1	1.59E-1	4.93E-2	1.95E-2	1.91E-2	4.67E-2	2.09E-2	2.12E-2	1.89E-2
	vr	**1.0**	**29.9**	**310**	**1991**	**2066**	**346**	**1733**	**1673**	**2103**

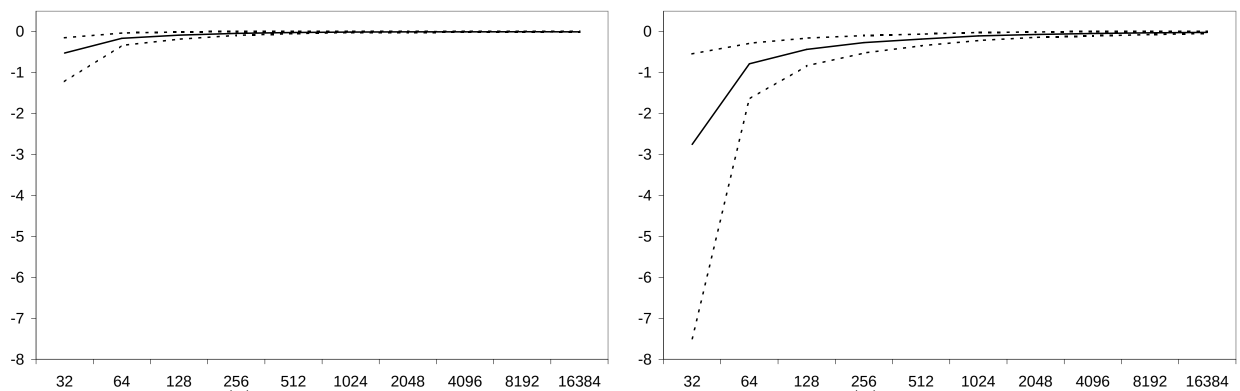
Table 7: Statistics for (WP^ν) as a function of ν, Lattice rule and Sobol with AV.

ν		LR+AV	SOB+AV	ν		LR+AV	SOB+AV
32	$\hat{\mu}$	-305.578	-306.069	1024	$\hat{\mu}$	-315.452	-315.482
	$\hat{\sigma}$	8.92E+0	7.35E+0		$\hat{\sigma}$	1.39E-1	1.37E-1
	vr	**27**	**40**		vr	**510**	**518**
64	$\hat{\mu}$	-312.915	-313.070	2048	$\hat{\mu}$	-315.496	-315.494
	$\hat{\sigma}$	2.82E+0	2.12E+0		$\hat{\sigma}$	5.46E-2	7.54E-2
	vr	**42**	**75**		vr	**1597**	**837**
128	$\hat{\mu}$	-314.719	-314.737	4096	$\hat{\mu}$	-315.498	-315.502
	$\hat{\sigma}$	7.23E-1	7.99E-1		$\hat{\sigma}$	3.95E-2	4.78E-2
	vr	**244**	**200**		vr	**1876**	**1285**
256	$\hat{\mu}$	-315.192	-315.211	8192	$\hat{\mu}$	-315.505	-315.506
	$\hat{\sigma}$	4.17E-1	4.50E-1		$\hat{\sigma}$	2.66E-2	2.47E-2
	vr	**302**	**259**		vr	**2156**	**2489**
512	$\hat{\mu}$	-315.366	-315.405	16384	$\hat{\mu}$	-315.505	-315.504
	$\hat{\sigma}$	2.49E-1	2.30E-1		$\hat{\sigma}$	1.36E-2	1.67E-2
	vr	**344**	**404**		vr	**4096**	**2719**



(a) Lattice rule. (b) Monte carlo.

Figure 5: Mean and 90% confidence interval for the optimal value in utility maximization problem with implicit constraints.



(a) Lattice rule. (b) Monte carlo.

Figure 6: Mean and 90% confidence interval for infeasibility $\min_i x_i$.

References

Z. Artstein and R. J.-B. Wets. Stability results for stochastic programs and sensors, allowing for discontinuous objective functions. *SIAM J. Optim.*, 4(3):537–550, 1994.

Z. Artstein and R. J.-B. Wets. Consistency of minimizers and the SLLN for stochastic programs. *J. Convex Anal.*, 2(1-2):1–17, 1995.

H. Attouch. *Variational convergence for functions and operators.* Pitman (Advanced Publishing Program), Boston, MA, 1984.

P. Billingsley. *Convergence of probability measures.* John Wiley & Sons Inc., New York, 2nd edition, 1999.

P. Boyle, M. Broadie, and P. Glasserman. Monte carlo methods for security pricing. *Journal of Economic Dynamics and Control*, 21:1267–1321, 1997.

P. Bratley, B. L. Fox, and L. E. Schrage. *A guide to simulation.* Springer-Verlag, 2 edition, 1987.

R. Caflisch, W. Morokoff, and A. Owen. Valuation of mortgage backed securities using brownian bridges to reduce effective dimension. *Journal of Computational Finance*, 1:27–46, 1997.

R. Cranley and T. N. L. Patterson. Randomization of number theoretic methods for multiple integration. *SIAM J. Numer. Anal.*, 13(6):904–914, 1976.

H. Faure. Discrépance de suites associées à un système de numération (en dimension s). *Acta Arith.*, 41(4):337–351, 1982.

B. L. Fox. Algorithm 647: Implementation and relative efficiency of quasirandom sequence generators. *ACM Transactions on Mathematical Software*, 12(4):362–376, 1986.

I. Friedel and A. Keller. Fast generation of randomized low discrepancy point sets. In K.-T. Fang, F. J. Hickernell, and H. Niederreiter, editors, *Monte Carlo and Quasi-Monte Carlo Methods 2000*, pages 257–273. Springer Verlag, 2002.

J. H. Halton. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer. Math.*, 2:84–90, 1960.

J. M. Hammersley. Monte Carlo methods for solving multivariable problems. *Ann. New York Acad. Sci.*, 86:844–874 (1960), 1960.

J. L. Higle. Variance reduction and objective function evaluation in stochastic linear programs. *Journal on Computing*, 10:236–247, 1998.

G. Infanger. Monte Carlo (importance) sampling within a Benders decomposition algorithm for stochastic linear programs. *Ann. Oper. Res.*, 39:69–95, 1992.

P. Jäckel. *Monte Carlo Methods in Finance.* John Wiley & Sons, 2002.

R. Kouwenberg. Scenario generation and stochastic programming models for asset liability management. *European J. Oper. Res.*, 134(2):279–292, 2001.

P. L'Ecuyer and C. Lemieux. Variance reduction via lattice rules. *Management Science*, 46(2): 1214–1235, 2000.

P. L'Ecuyer and C. Lemieux. Recent advances in randomized quasi-monte carlo methods. In M. Dror, P. L'Ecuyer, and F. Szidarovszki, editors, *Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications*, pages 419–474. Kluwer Academic Publishers, 2002.

J. Linderoth, A. Shapiro, and S. Wright. The empirical behavior of sampling methods for stochastic programming. Optimization technical report 02-01, Computer Sciences Department, University of Wisconsin-Madison, 2002.

R. Lucchetti, G. Salinetti, and R. J.-B. Wets. Uniform convergence of probability measures: topological criteria. *J. Multivariate Anal.*, 51(2):252–264, 1994.

R. Lucchetti and R. J.-B. Wets. Convergence of minima of integral functionals, with applications to optimal control and stochastic optimization. *Statist. Decisions*, 11(1):69–84, 1993.

M. Matsumoto and T. Nishimura. Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions Modeling Computer Simulation*, 8(1):3–30, 1998.

H. Niederreiter. Low-discrepancy and low-dispersion sequences. *J. Number Th.*, 30:51–70, 1988.

H. Niederreiter. *Random number generation and quasi-Monte Carlo methods*, volume 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992.

T. Pennanen and M. Koivu. Epi-convergent discretizations of stochastic programs via integration quadratures. *Stochastic Programming E-print Series*, 2003.

W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical recipes in C, The art of scientific computing*. Cambridge University Press, Cambridge, 2nd edition, 1992.

R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.

A. Shapiro. Stochastic programming by monte carlo simulation methods. *Stochastic Programming E-Print Series*, 2000.

A. Shapiro. Inference of statistical bounds for multistage stochastic programming problems. *Mathematical Methods of Operations Research*, 58:57–68, 2003.

I. H. Sloan and S. Joe. *Lattice methods for multiple integration*. Oxford Science Publications. The Clarendon Press Oxford University Press, New York, 1994.

I. M. Sobol'. The distribution of points in a cube and the approximate evaluation of integrals. *U.S.S.R. Computational Math. And Math. Phys.*, (4):86–112, 1967.

B. Tuffin. On the use of low discrepancy sequences in Monte Carlo methods. *Monte Carlo Methods and Appl.*, 2(4):295–320, 1996.

X. Wang and K. T. Fang. The effective dimension and quasi-monte carlo integration. *Journal of Complexity*, 19:101–124, 2002.

M. Zervos. On the epiconvergence of stochastic optimization problems. *Math. Oper. Res.*, 24(2): 495–508, 1999.